International Journal of Science Engineering and Advance Technology

Machine Learning Text Categorization In OSN To Filter Unwanted Messages

Kota Prudhvee Raj[1] ,Kotaru Anil Chowdary[2]
[1]*M.Tech, Department of Computer Science and Engineering,*
Sir C R Reddy College of Engineering, Eluru, West Godavari District
Andhra Pradesh, India – 534007
[2]*Associate Professor, Department of Computer Science and Engineering,*
Sir C R Reddy College of Engineering, Eluru, West Godavari District
Andhra Pradesh, India – 534007
Prudhvi9889@gmail.com, anilchow@gmail.com

Abstract—One fundamental issue in today's Online Social Networks (OSNs) is to give users the ability to control the messages posted on their own private space to avoid that unwanted content is displayed. Up to now, OSNs provide little support to this requirement. To fill the gap, in this paper, we propose a system allowing OSN users to have a direct control on the messages posted on their walls. This is achieved through a flexible rule-based system, that allows users to customize the filtering criteria to be applied to their walls, and a Machine Learning-based soft classifier automatically labeling messages in support of content-based filtering.

Keywords: Online social networks, information filtering, short text classification, policy-based personalization**.**

Introduction

Today's modern life is totally based on Internet. Now a days people cannot imagine life without Internet. Also, OSNs are just a part of modern life. From last few years people share their views, ideas, information with each other using social networking sites. Such communications may involve different types of contents like text, image, audio and video data. But, in today's OSN , there is a very high chance of posting unwanted content on particular public/private areas, called in general walls. So, to control this type of activity and prevent the unwanted messages which are written on user's wall we can implement filtering rules (FR) in our system. Also, Black List (BL) will maintain in this system .We present this system as www.winow.in on the internet. It can be used to give users the ability to automatically control the messages written on their own walls, by filtering out unwanted messages. The huge and dynamic character of these data creates the premise for the employment of web content mining strategies aimed to automatically discover useful information dormant within the data.

OSNs provide support to prevent unwanted messages on user walls. For example, Face book allows users to state who is allowed to insert messages in their walls (i.e., friends, friends of friends, or defined groups of friends). However, no content-based preferences are supported and therefore it is not possible to prevent undesired messages, such as political or vulgar ones, no matter of the user who posts them.

Providing this service is not only a matter of using previously defined web content mining techniques for a different application, rather it requires to design ad hoc classification strategies. This is because wall messages are constituted by short text for which traditional classification methods have serious limitations since short texts do not provide sufficient word occurrences.

2. Related Work

In www.winow.in information filtering techniques are used to remove unwanted contents by using customizable content based filtering rules, Machine learning approach; according to user's interest and recommends an item.

Recommender systems works in following ways

- Content based filtering
- Collaborative filtering
- Policy based filtering

A. Content-based filtering

In content based filtering to check the user's interest and previous activity as well as item uses by users best match is found [10]. For example OSNs such as Face book, Orkut used content based filtering policy. In that by checking users profile attributes like education, work area, hobbies etc. suggested friend request may send. The main purpose of content based filtering, the system is able to learn from user's actions related to a particular content source and use them for other content types.

B. ***Collaborative filtering***

In collaborative filtering information will be selected on the basis of user's preferences, actions, predicts, likes, and dislikes. Match all this information with other users to find out similar items. Large dataset is required for collaborative filtering system. According to user's likes and dislikes items are rated.

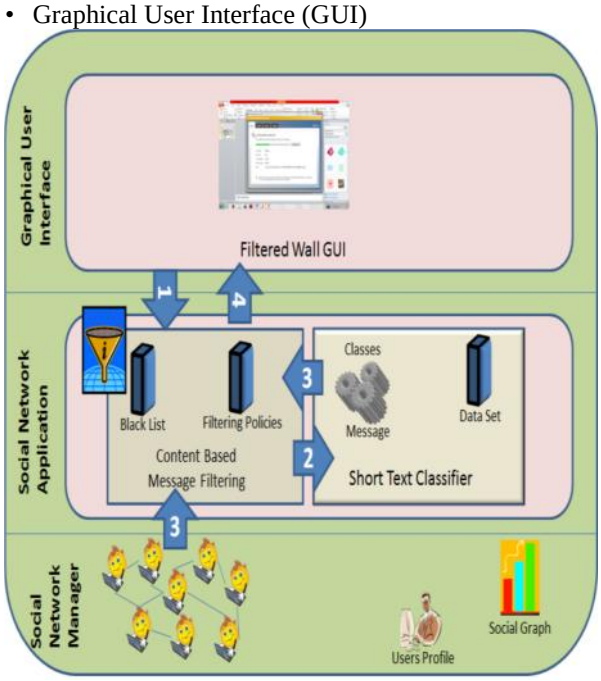
C. ***Policy-based filtering***

In policy based filtering system users filtering ability is represented to filter wall messages according to filtering criteria of the user. Twitter is the best example for policy based filtering.[1] In that communication policy can be defines between two communicating parties.

3. Filtered Wall Architecture

Three Tier architecture is used in OSN services. These three layers are

- Social Network Manager (SNM)
- Social Network Application (SNA)
- Graphical User Interface (GUI)



1. Social Network Manager (SNM)

The initial layer is Social Network Manager layer provides the essential OSN functionalities (i.e., profile and relationship administration).It also maintains all the data regarding to the user profile.[2] After maintaining and administrating all users data will provide for second layer for applying Filtering Rules (FR) and Black lists (BL).

2. Social Network Application (SNA)

In second layer Content Based Message Filtering (CBMF) and Short Text Classifier is composed. This is very important layer for the message categorization according to its CBMF filters. Also Black list is maintained for the user who sends frequently bad words in message.

3. Graphical User Interface (GUI)

Third layer provides Graphical User Interface to the user who wants to post his messages as a input. In this layer Filtering Rules (FR) are used to filter the unwanted messages and provide Black list (BL) for the user who are temporally prevented to publish messages on user's wall.

The GUI also consists of Filtered Wall (FW) where the user is able to see his desirable messages.[5] As shown in Fig. 1 points summarized as follows:

1. After entering the private wall of one of his/her associates, the user attempts to post a message, which is captured by FW.
2. A ML-based text classifier extracts metadata from the content of the message.
3. FW uses metadata provided by the classifier, mutually with data extorted from the social graph and users' profiles, to implement the filtering and BL rules.
4. Depending on the result of the previous step, the message will be available or filtered by FW.

4. Mathematical Model

A. For Filtering Rules:

1) Input

Filtering Rules are customizable by the user. User can have authority to decide what contents should be blocked or displayed on his wall by using Filtering rules. For specify a Filtering rules user profile as well as user social relationship will be considered.

FR= {Actor, UserSpec, ContentSpec}

FR is dependent on following factors

- Author
- UserSpec
- ContentSpec
- Action

Author is a person who defines the rules.

UserSpec denotes the set of OSN user.

ContentSpec is a Boolean expression defined on content.

2) Process

FM={UserSpec,contentSpec==category(Violence,Vul
gar,offensive,Hate,Sextual)}

- FM
- UserSpec
- ContentSpec

Here,

FM Block Messages in percentage

UserSpec Denotes set of users

ContentSpec Category of specified contents in message.

In processing, after giving input message, the system will compare the text with the different categories which are prevented. If message found in that prevented type of category then message will display to the user that "can't send this type of messages."

Process denotes the action to be performed by the system on the messages matching Content-Spec and created by users identified by UserSpec.

3) Output

PFM= {ContentSpec, M||Y}

• *PFM* Percentages of filtered message in a year or month.

In general, more than a filtering rule can apply to the same user. A message is therefore published only if it is not blocked by any of the filtering rules that apply to the message creator.

B. Blacklists

BLs are directly managed by the system, This should be able to determine who are the users to be inserted in the BL and decide when users' retention in the BL is finished. To enhance flexibility, such information is given to the system through a set of rules, hereafter called BL rules.

Definition 3 (BL rule).

1) Input

INPUT = {Actor, UserSpec, UserBehavior} Where

- author is the OSN user who specifies the rule, i.e., the wall owner;
- UserSpec is a creator specification, specified according to Definition 1;
- UserBehavior consists of Message sending category of User.
-

2) Process

BL={UserSpec,ContentSpec,T}

- UserSpec
- ContentSpec
- T

UserSpec Creator Specification

ContentSpec Message send by User.

T Messages is the total number of messages that each OSN user sent.

3) Output

BL={UserSpec,ContentSpec,T>3 ,P}

- UserSpec
- ContentSpec
- T > 3

UserSpec Creator Specification

ContentSpec Message send by User.

T Prevented Message count is greater than 3 times then Messagecreator will put into Black list automatically for specific time period **P.**

5. Online Setup Assistant For Frs Thresholds

OSA presents the user with a set of messages selected from the data set. For each message, the user tells the system the decision to accept or reject the message. The collection and processing of user decisions on an adequate set of messages distributed over all the classes allows to compute customized thresholds representing the user attitude in accepting or rejecting certain contents.

6. Dataset

Facebook and Twitter are two representative OSNs. We use data collected from both sites in the study. The Facebook dataset contains 187 million wall posts generated by roughly 3.5 million users in total, between January of 2008 and June of 2009 [30]. For the Twitter data collection, we first download trending topics, *i.e.* popular topics, from the website *What the Trend* [2], which provides a regularly updated list of trending topics. We then download from Twitter all public tweets that contain the trending topics while the topics are still popular via Twitter APIs. For example, while the topic "MLB" is trending, we keep downloading all tweets that contain the word "MLB" from Twitter. For each tweet we also obtain the user ID that generates it along with its friend number, *i.e.* the number of users it follows. The Twitter dataset contains over 17 million tweets related to trending topics that were generated between June 1, 2011 and July 21, 2011. The primary form of communication in Facebook and Twitter is called "wall post" and "tweet", respectively. From now on, we use the term "message" to refer to both of them for the ease of description.

One thousand two hundred and sixty-six messages from publicly accessible Italian groups have been selected and extracted by means of an automated procedure that removes undesired spam messages and, for each message, stores the message body and the name of the group from which it originates. The messages come from the group's webpage section, where any registered user can post a new message or reply to messages already posted by other users. The set of classes considered in our experiments is Ω= {Neutral, Violence, Vulgar, Offensive, Hate, Sex}, where Ω- {Neutral} are the second-level classes. Total 31 percentages belongs to the elements D for the Neutral class.

7. Application

This application is useful for common people who don't want to write any unwanted messages like vulgar, political, sexual messages on his\her own wall by any third person.

Mostly, this type of activities are happen with some famous personalities, So if this facility will

provide with OSN sites then people can protect his wall from this type of malpractices.

8. Conclusion

Existing system is used to filter undesired messages from OSNs wall using customizable filtering rules (FR) enhancing through Black lists (BLs). In present system (www.winow.in), we are more focus on an investigation of two interdependent tasks in depth. This system approach decides when user should be inserted into a black list.

The system developed GUI and a set of tools which make BLs and FRs specifications more simple and easy. Investigation tools may be able to automatically recommend trust value of the user. The primary work of this system is to find out trust values used for OSN access control. In this system we will provide only core set of functionalities which are available in current OSNs

like Facebook, Orkut, Twitter,etc. In existing OSNs have some difficulties in understanding to the average users regarding privacy settings. But this problem will be overcome in present OSNs system.

References

[1] Marco Vanetti, Elisabetta Binaghi, Elena Ferrari, Barbara Carminati, and Moreno Carullo," A System to Filter Unwanted Messages from OSN User Walls" IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 25, NO. 2, FEBRUARY 2013.

[2] A. Adomavicius and G. Tuzhilin, "Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions," IEEE Trans. Knowledge and Data Eng., vol. 17, no. 6, pp. 734-749, June 2005.

[3] M. Chau and H. Chen, "A Machine Learning Approach to Web Page Filtering Using Content and Structure Analysis," Decision Support Systems, vol. 44, no. 2, pp. 482-494, 2008.

[4] M. Vanetti, E. Binaghi, B. Carminati, M. Carullo, and E. Ferrari, "Content-Based Filtering in On-Line Social Networks," Proc. ECML/PKDD Workshop Privacy and Security Issues in Data
Mining and Machine Learning (PSDML '10), 2010.

[5] F. Sebastiani, "Machine Learning in Automated Text Categorization," ACM Computing Surveys, vol. 34, no. 1, pp. 1- 47, 2002.